Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Journal of Multivariate Analysis 97 (2006) 1815 – 1828

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

A class of proper priors for Bayesian simultaneous prediction of independent Poisson observables

Fumiyasu Komaki*

Department of Mathematical Informatics, Graduate School of Information Science and Technology, the University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Received 30 November 2004
Available online 23 January 2006

Abstract

Simultaneous prediction and parameter inference for the independent Poisson observables model are considered. A class of proper prior distributions for Poisson means is introduced. Bayesian predictive densities and estimators based on priors in the introduced class dominate the Bayesian predictive density and estimator based on the Jeffreys prior under Kullback–Leibler loss.

AMS 1991 subject classification: 62F15; 62C10

Keywords: Jeffreys prior; Kullback–Leibler divergence; Predictive density; Shrinkage prior

1. Introduction

Suppose that $x = (x_1, x_2, \dots, x_d)$ and $y = (y_1, y_2, \dots, y_d)$ are sets of independent Poisson random variables with mean vectors $(a\lambda_1, a\lambda_2, \dots, a\lambda_d)$ and $(b\lambda_1, b\lambda_2, \dots, b\lambda_d)$, respectively. Here, $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_d)$ is an unknown d-dimensional parameter and a and b are known positive real numbers. The probability densities of x and y are given by

$$p(x|\lambda) = \prod_{i=1}^{d} p(x_i|\lambda) = \exp\{-(a\lambda_1 + a\lambda_2 + \cdots + a\lambda_d)\} \frac{(a\lambda_1)^{x_1}}{x_1!} \frac{(a\lambda_2)^{x_2}}{x_2!} \cdots \frac{(a\lambda_d)^{x_d}}{x_d!} \tag{1}$$

* Fax: +81 3 5841 8592.
E-mail address: komaki@mist.i.u-tokyo.ac.jp.

0047-259X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2005.12.002

and

$$p(y|\lambda) = \prod_{i=1}^{d} p(y_i|\lambda) = \exp\{-(b\lambda_1 + b\lambda_2 + \cdots + b\lambda_d)\} \frac{(b\lambda_1)^{y_1}}{y_1!} \frac{(b\lambda_2)^{y_2}}{y_2!} \cdots \frac{(b\lambda_d)^{y_d}}{y_d!}, \tag{2}$$

respectively. In the following, we call the model defined by (1) and (2) as the independent Poisson observables model.

We consider the problem of constructing a predictive density $\hat{p}(y; x)$ for the unobserved variable y by using the observed variable x. The Kullback–Leibler divergence

$$D(p(y|\lambda), \hat{p}(y; x)) = \sum_{y} p(y|\lambda) \log \frac{p(y|\lambda)}{\hat{p}(y; x)}$$

from the true density $p(y|\lambda)$ to a predictive density $\hat{p}(y; x)$ is adopted as a loss function. Predictive distribution theory can be regarded as a natural generalization of estimation theory under Kullback–Leibler loss since the Kullback–Leibler loss for a plug-in density $p(y|\hat{\lambda}(x))$ can be regarded as a loss for the estimator $\hat{\lambda}(x)$.

Non-informative prior densities or vague prior densities are often used to construct Bayesian predictive densities. The Jeffreys prior naturally arises as a non-informative prior from various discussions based on the Kullback–Leibler divergence. However, Bayesian methods based on the Jeffreys prior do not always perform satisfactorily especially in problems with multidimensional parameters [3].

Recently, several studies on the use of shrinkage priors for prediction have been carried out. Shrinkage priors give more weight to parameter values close to a point or a subspace in the parameter space than the Jeffreys prior does.

For the d-dimensional normal model $N_d(\mu, I)$ $(d \geqslant 3)$, where μ is an unknown vector and I is the identity matrix, Komaki [4] showed that the Bayesian predictive density based on the improper shrinkage prior $\pi_{\mathrm{S}}(\mu) \propto \|\mu\|^{-(d-2)}$ introduced by Stein [8] dominates the Bayesian predictive density based on the Jeffreys prior $\pi_{\mathrm{I}}(\mu) \propto 1$, which is the best predictive density invariant under the translation group. Liang [7] showed that there exist Bayesian predictive densities for $N_d(\mu, I)$ based on proper priors dominating that based on the Jeffreys prior. George et al. [1] obtained several sufficient conditions for prior densities under which a Bayesian predictive density for $N_d(\mu, I)$ based on a prior dominates the Bayesian predictive density based on the Jeffreys prior. They showed that a class of proper priors for the d-dimensional normal model $N_d(\mu, I)$ $(d \geqslant 5)$ introduced by Strawderman [9] satisfies the sufficient conditions.

For the independent Poisson observables model, Komaki [5] introduced an improper shrinkage prior. The Bayesian predictive density based on the prior is admissible under Kullback–Leibler loss and dominates the Bayesian predictive density based on the Jeffreys prior.

In the present paper, we introduce a class of priors for the independent Poisson observables model that includes proper priors when $d \geqslant 5$. It is shown that Bayesian predictive densities based on a prior in the introduced class dominate the Bayesian predictive density based on the prior

$$\tilde{\pi}_{\beta}(\lambda) := \prod_{i=1}^{d} \lambda_i^{\beta_i - 1}$$

with $\beta_i > 0$ $(i = 1, 2, \ldots, d)$. When $\beta_i = 1/2$ $(i = 1, 2, \ldots, d)$, the prior $\tilde{\pi}_{\beta}(\lambda)$ coincides with the Jeffreys prior.

In Section 2, a class of priors for the independent Poisson observables model is introduced. When $d \geqslant 5$, the class of priors includes proper ones. It is shown that Bayesian predictive densities based on a prior in the introduced class dominate the Bayesian predictive density based on the prior $\tilde{\pi}_\beta(\lambda)$. In Section 3, it is shown that (possibly generalized) Bayes estimators based on a prior in the introduced class dominate the generalized Bayes estimator based on the prior $\tilde{\pi}_\beta(\lambda)$. In Section 4, some discussions from an asymptotic viewpoint are given.

2. A class of prior densities

We introduce a class of prior densities

$$\pi_{\beta,c,\kappa}(\lambda) := \left(\prod_{i=1}^{d} \lambda_i^{\beta_i - 1}\right) \int_0^\infty \exp\left(-\frac{\sum_{j=1}^d \lambda_j}{s}\right) s^{-\sum_{k=1}^d \beta_k} (s+\kappa)^{-(c+1)}\, \mathrm{d}s, \tag{3}$$

where $\sum_{i=1}^d \beta_i + c > 0$, $\kappa > 0$, and $\beta_i > 0$ $(i = 1, 2, \ldots, d)$. We ignore normalizing constants for prior densities since the results are unaffected. The prior density $\pi_{\beta,c,\kappa}(\lambda)$ is a scale mixture of products $\prod_{i=1}^d \{\lambda_i^{\beta_i - 1} s^{-\beta_i} \exp(-\lambda_i/s)\}$ of Gamma densities. When $c > 0$, $\pi_{\beta,c,\kappa}(\lambda)$ is a proper prior density. When $c = -1$, $\pi_{\beta,c,\kappa}(\lambda)$ coincides with an improper shrinkage prior density

$$\pi_\beta(\lambda) := \left(\sum_{i=1}^d \lambda_i\right)^{-\sum_{j=1}^d \beta_j + 1} \prod_{k=1}^d \lambda_k^{\beta_k - 1}$$

investigated by Komaki [5]. In the limit $\kappa \to \infty$, $\pi_{\beta,c,\kappa}(\lambda)$ also coincides with $\pi_\beta(\lambda)$.

Theorem 1. *The Bayesian predictive density based on the prior defined by* (3) *with* $\sum_{i=1}^d \beta_i + c > 0$, $\kappa > 0$, *and* $\beta_i > 0$ $(i = 1, 2, \ldots, d)$ *is given by*

$$p_{\pi_{\beta,c,\kappa}}(y|x) = \frac{a^{\sum_i x_i + 1} b^{\sum_j y_j}}{(a+b)^{\sum_k x_k + \sum_l y_l + 1}} \times \frac{\displaystyle\int_0^1 r^{\sum_i x_i} (1-r)^{\sum_j \beta_j + c - 1} \left[1 - \left\{1 - \frac{1}{(a+b)\kappa}\right\} r\right]^{-(c+1)} \mathrm{d}r}{\displaystyle\int_0^1 \bar{r}^{\sum_k x_k} (1-\bar{r})^{\sum_l \beta_l + c - 1} \left\{1 - \left(1 - \frac{1}{a\kappa}\right) \bar{r}\right\}^{-(c+1)} \mathrm{d}\bar{r}\; y_1! y_2! \cdots y_d!}.$$

Proof. By using Lemma 1 in the Appendix, we have

$$p_{\pi_{\beta,c,\kappa}}(y|x) = \frac{\displaystyle\int \pi_{\beta,c,\kappa}(\lambda) \prod_{i=1}^d \left\{\exp(-a\lambda_i) \frac{(a\lambda_i)^{x_i}}{x_i!}\right\} \prod_{j=1}^d \left\{\exp(-b\lambda_j) \frac{(b\lambda_j)^{y_j}}{y_j!}\right\} \mathrm{d}\lambda}{\displaystyle\int \pi_{\beta,c,\kappa}(\bar{\lambda}) \prod_{k=1}^d \left\{\exp(-a\bar{\lambda}_k) \frac{(a\bar{\lambda}_k)^{x_k}}{x_k!}\right\} \mathrm{d}\bar{\lambda}}$$

$$
\begin{aligned}
&= \frac{\int \pi_{\beta,c,\kappa}(\lambda) \prod_{i=1}^d \left[\exp\{-(a+b)\lambda_i\}\{(a+b)\lambda_i\}^{x_i+y_i} \right] \mathrm{d}\lambda}{\int \pi_{\beta,c,\kappa}(\bar{\lambda}) \prod_{k=1}^d \left\{ \exp(-a\bar{\lambda}_k)(a\bar{\lambda}_k)^{x_k} \right\} \mathrm{d}\bar{\lambda}} \prod_{j=1}^d \frac{a^{x_j} b^{y_j}}{(a+b)^{x_j+y_j} y_j!} \\
&= \frac{a^{\sum_i x_i+1} b^{\sum_j y_j}}{(a+b)^{\sum_k x_k + \sum_l y_l + 1}} \\
&\times \frac{\kappa^{-(c+1)} \prod_{i=1}^d \Gamma(x_i+\beta_i) \int_0^1 r^{\sum_j x_j} (1-r)^{\sum_k \beta_k + c-1} \left[1 - \left\{ 1 - \frac{1}{(a+b)\kappa} \right\} r \right]^{-(c+1)} \mathrm{d}r}{\kappa^{-(c+1)} \prod_{l=1}^d \Gamma(x_l+\beta_l) \int_0^1 \bar{r}^{\sum_m x_m} (1-\bar{r})^{\sum_n \beta_n + c-1} \left\{ 1 - \left(1 - \frac{1}{a\kappa} \right) \bar{r} \right\}^{-(c+1)} \mathrm{d}\bar{r}\; y_1! y_2! \cdots y_d!}.
\end{aligned}
$$

Thus we obtain the desired result. □

Theorem 2. *Suppose that $\pi_{\beta,c,\kappa}(\lambda)$ is a prior density defined by* (3) *with $\sum_{i=1}^d \beta_i + c > 0, \kappa > 0$, and $\beta_i > 0\, (i = 1, 2, \ldots, d)$. If $c > -1, \sum_{i=1}^d \beta_i - c - 2 \geqslant 0$, and $a \geqslant 1/\kappa$, the Bayesian predictive density based on the prior $\pi_{\beta,c,\kappa}(\lambda)$ dominates that based on the prior $\tilde{\pi}_\beta(\lambda) := \prod_{i=1}^d \lambda_i^{\beta_i - 1}$.*

Proof. The Bayesian predictive density based on a prior density $\pi(\lambda)$ is given by

$$
p_\pi(y|x) = \frac{\int \pi(\lambda) \prod_{i=1}^d \left[\exp\{-(a+b)\lambda_i\}\{(a+b)\lambda_i\}^{x_i+y_i} \right] \mathrm{d}\lambda}{\int \pi(\bar{\lambda}) \prod_{k=1}^d \left\{ \exp(-a\bar{\lambda}_k)(a\bar{\lambda}_k)^{x_k} \right\} \mathrm{d}\bar{\lambda}} \prod_{j=1}^d \frac{a^{x_j} b^{y_j}}{(a+b)^{x_j+y_j} y_j!}.
$$

The difference between the risk functions of Bayesian predictive densities based on $\pi_1(\lambda)$ and $\pi_2(\lambda)$ is

$$
\begin{aligned}
&E[D(p(y|\mu), p_{\pi_1}(y|x))|\lambda] - E[D(p(y|\mu), p_{\pi_2}(y|x))|\lambda] \\
&\quad = E\left[\sum_y p(y|\lambda) \log \frac{p_{\pi_2}(y|x)}{p_{\pi_1}(y|x)} \middle| \lambda \right] \\
&\quad = E\left[\log \int \pi_2(\bar{\lambda}) \prod_{i=1}^d \left(\exp\{-(a+b)\bar{\lambda}_i\}\{(a+b)\bar{\lambda}_i\}^{x_i+y_i} \right) \mathrm{d}\bar{\lambda} \middle| \lambda \right] \\
&\quad\quad - E\left[\log \int \pi_2(\bar{\lambda}) \prod_{j=1}^d \left\{ \exp(-a\bar{\lambda}_j)(a\bar{\lambda}_j)^{x_j} \right\} \mathrm{d}\bar{\lambda} \middle| \lambda \right] \\
&\quad\quad - E\left[\log \int \pi_1(\bar{\lambda}) \prod_{k=1}^d \left(\exp\{-(a+b)\bar{\lambda}_k\}\{(a+b)\bar{\lambda}_i\}^{x_k+y_k} \right) \mathrm{d}\bar{\lambda} \middle| \lambda \right] \\
&\quad\quad + E\left[\log \int \pi_1(\bar{\lambda}) \prod_{l=1}^d \left\{ \exp(-a\bar{\lambda}_l)(a\bar{\lambda}_l)^{x_l} \right\} \mathrm{d}\bar{\lambda} \middle| \lambda \right].
\end{aligned}
$$

Therefore, if

$$E\left[\log \int \pi_2(\bar{\lambda}) \prod_{i=1}^{d} \left\{ \exp(-t\bar{\lambda}_i)(t\bar{\lambda}_i)^{x_i} \right\} \mathrm{d}\bar{\lambda} \middle| \lambda \right] \\ -E\left[\log \int \pi_1(\bar{\lambda}) \prod_{j=1}^{d} \left\{ \exp(-t\bar{\lambda}_j)(t\bar{\lambda}_j)^{x_j} \right\} \mathrm{d}\bar{\lambda} \middle| \lambda \right], \tag{4}$$

where x_i $(i = 1, \ldots, d)$ is a Poisson random variable with mean $t\lambda_i$, is a strictly increasing function of $t \geqslant \tau$, where τ is a fixed nonnegative constant, the Bayesian predictive density based on $\pi_2(\lambda)$ dominates that based on $\pi_1(\lambda)$ when $a \geqslant \tau$.

From Lemma 1, we have

$$\log \int \pi_{\beta,c,\kappa}(\lambda) \prod_{k=1}^{d} \left\{ \exp(-t\lambda_k)(t\lambda_k)^{x_k} \right\} \mathrm{d}\lambda_1\, \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_d \\ = -\log t - (c+1)\log\kappa + \sum_{i=1}^{d} \log\Gamma(x_i + \beta_i) \\ + \log \int_0^1 r^{\sum_i x_i}(1-r)^{\sum_j \beta_j + c - 1} \left\{ 1 - \left(1 - \frac{1}{t\kappa}\right) r \right\}^{-(c+1)} \mathrm{d}r.$$

Since

$$\log \int \tilde{\pi}_\beta(\lambda) \prod_{i=1}^{d} \left\{ \exp(-t\lambda_i)(t\lambda_i)^{x_i} \right\} \mathrm{d}\lambda_1\, \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_d = -\sum_{i=1}^{d} \beta_i \log t + \sum_{j=1}^{d} \log\Gamma(x_j + \beta_j),$$

we have

$$E\left[\log \int \pi_{\beta,c,\kappa}(\bar{\lambda}) \prod_{k=1}^{d} \left\{ \exp(-t\bar{\lambda}_k)(t\bar{\lambda}_k)^{x_k} \right\} \mathrm{d}\bar{\lambda} \middle| \lambda \right] \\ -E\left[\log \int \tilde{\pi}_\beta(\bar{\lambda}) \prod_{k=1}^{d} \left\{ \exp(-t\bar{\lambda}_k)(t\bar{\lambda}_k)^{x_k} \right\} \mathrm{d}\bar{\lambda} \middle| \lambda \right] \\ = \left(\sum_i \beta_i - 1 \right) \log t - (c+1)\log\kappa \\ + E\left[\log \int_0^1 r^{\sum_j x_j}(1-r)^{\sum_k \beta_k + c - 1} \left\{ 1 - \left(1 - \frac{1}{t\kappa}\right) r \right\}^{-(c+1)} \mathrm{d}r \middle| \lambda \right]. \tag{5}$$

Here, $\sum_j x_j$ is a Poisson random variable with mean $t\mu$, where $\mu := \sum_i \lambda_i$. When $\mu = 0$, (5) is obviously a strictly increasing function of $t > 0$ since $\sum_i \beta_i - 1 \geqslant c + 1 > 0$. In the following, we assume $\mu > 0$ and show that (5) is a strictly increasing function of $t \geqslant 1/\kappa$.

Let $z_t = 1 - 1/(t\kappa)$. The derivative of (5) is given by

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\Bigg[\left(\sum_i \beta_i - 1\right)\log t - (c+1)\log\kappa \\
&\quad + \sum_{n=0}^{\infty}\exp(-t\mu)\frac{(t\mu)^n}{n!}\log\int_0^1 r^n(1-r)^{\sum_i \beta_i + c - 1}(1-z_t r)^{-(c+1)}\,\mathrm{d}r\Bigg] \\
&= \left(\sum_i \beta_i - 1\right)\frac{1}{t} + \frac{1}{t}\sum_{n=0}^{\infty}(-t\mu)\exp(-t\mu)\frac{(t\mu)^n}{n!}\log\int_0^1 r^n(1-r)^{\sum_i \beta_i + c - 1}(1-z_t r)^{-(c+1)}\,\mathrm{d}r \\
&\quad + \frac{1}{t}\sum_{n=0}^{\infty}(t\mu)\exp(-t\mu)\frac{(t\mu)^n}{n!}\log\int_0^1 r^{n+1}(1-r)^{\sum_i \beta_i + c - 1}(1-z_t r)^{-(c+1)}\,\mathrm{d}r \\
&\quad + \frac{c+1}{t^2\kappa}\sum_{n=0}^{\infty}\exp(-t\mu)\frac{(t\mu)^n}{n!}\frac{\int_0^1 r^{n+1}(1-r)^{\sum_i \beta_i + c - 1}(1-z_t r)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^n(1-r)^{\sum_j \beta_j + c - 1}(1-z_t r)^{-(c+1)}\,\mathrm{d}r} \\
&= \frac{1}{t}\Bigg[\sum_i \beta_i - 1 + \sum_{n=1}^{\infty}\exp(-t\mu)\frac{(t\mu)^n}{n!}n\log\frac{\int_0^1 r^n(1-r)^{\sum_i \beta_i + c - 1}(1-z_t r)^{-(c+1)}\,\mathrm{d}r}{\int_0^1 r^{n-1}(1-r)^{\sum_j \beta_j + c - 1}(1-z_t r)^{-(c+1)}\,\mathrm{d}r} \\
&\quad + (c+1)(1-z_t)\sum_{n=0}^{\infty}\exp(-t\mu)\frac{(t\mu)^n}{n!}\frac{\int_0^1 r^{n+1}(1-r)^{\sum_i \beta_i + c - 1}(1-z_t r)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^n(1-r)^{\sum_j \beta_j + c - 1}(1-z_t r)^{-(c+1)}\,\mathrm{d}r}\Bigg] \\
&= \frac{1}{t}\Bigg[\exp(-t\mu)\left\{\sum_i \beta_i - 1 + (c+1)(1-z_t)\frac{\int_0^1 r(1-r)^{\sum_i \beta_i + c - 1}(1-z_t r)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 (1-r)^{\sum_j \beta_j + c - 1}(1-z_t r)^{-(c+1)}\,\mathrm{d}r}\right\} \\
&\quad + \sum_{n=1}^{\infty}\exp(-t\mu)\frac{(t\mu)^n}{n!}\left\{\sum_i \beta_i - 1 + n\log\frac{\int_0^1 r^n(1-r)^{\sum_i \beta_i + c - 1}(1-z_t r)^{-(c+1)}\,\mathrm{d}r}{\int_0^1 r^{n-1}(1-r)^{\sum_j \beta_j + c - 1}(1-z_t r)^{-(c+1)}\,\mathrm{d}r}\right. \\
&\quad \left. + (c+1)(1-z_t)\frac{\int_0^1 r^{n+1}(1-r)^{\sum_i \beta_i + c - 1}(1-z_t r)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^n(1-r)^{\sum_j \beta_j + c - 1}(1-z_t r)^{-(c+1)}\,\mathrm{d}r}\right\}\Bigg].
\end{aligned}
$$

Since $\sum_i \beta_i - 1 > 0$, $c + 1 > 0$, and $0 \leqslant z_t < 1$, the first term in the square brackets is positive. Therefore, it is enough to show that $\psi(n, \beta, c, z)$ defined below is nonnegative for every positive integer n to obtain the desired result. In the following we abbreviate z_t to z.

$$
\begin{aligned}
\psi(n,\beta,c,z) &:= \sum_i \beta_i - 1 + n\log\frac{\int_0^1 r^n(1-r)^{\sum_i \beta_i + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r}{\int_0^1 r^{n-1}(1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r} \\
&\quad + (c+1)(1-z)\frac{\int_0^1 r^{n+1}(1-r)^{\sum_i \beta_i + c - 1}(1-zr)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^n(1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r} \\
&= \sum_i \beta_i - 1 + (c+1)(1-z)\frac{\int_0^1 r^{n+1}(1-r)^{\sum_i \beta_i + c - 1}(1-zr)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^n(1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r} \\
&\quad - n\log\frac{n + \sum_i \beta_i + c}{n} \\
&\quad - n\log\left\{\frac{n}{n + \sum_i \beta_i + c}\frac{\int_0^1 r^{n-1}(1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r}{\int_0^1 r^n(1-r)^{\sum_k \beta_k + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\geqslant \sum_i \beta_i - 1 - n \log \frac{n + \sum_i \beta_i + c}{n} \\
&\quad + (c+1)(1-z) \frac{\int_0^1 r^{n+1}(1-r)^{\sum_i \beta_i + c - 1}(1-zr)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^{n}(1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r} \\
&\quad - \frac{n^2}{n + \sum_i \beta_i + c} \frac{\int_0^1 r^{n-1}(1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r}{\int_0^1 r^{n}(1-r)^{\sum_k \beta_k + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r} + n \\
&= \sum_i \beta_i - 1 - n \log \frac{n + \sum_i \beta_i + c}{n} \\
&\quad + (c+1)(1-z) \frac{\int_0^1 r^{n+1}(1-r)^{\sum_i \beta_i + c - 1}(1-zr)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^{n}(1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r} \\
&\quad - \frac{n^2}{n + \sum_i \beta_i + c} \frac{\int_0^1 r^{n-1}(1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r}{\int_0^1 r^{n}(1-r)^{\sum_k \beta_k + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r} \\
&\quad + \frac{n(n + \sum_i \beta_i - 1)}{n + \sum_j \beta_j + c} + \frac{n(c+1)}{n + \sum_i \beta_i + c} \\
&= \left(\sum_i \beta_i + c \right) \frac{n + \sum_j \beta_j - 1}{n + \sum_k \beta_k + c} - n \log \frac{n + \sum_i \beta_i + c}{n} \\
&\quad + (c+1)(1-z) \frac{\int_0^1 r^{n+1}(1-r)^{\sum_i \beta_i + c - 1}(1-zr)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^{n}(1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r} \\
&\quad - \frac{n}{n + \sum_i \beta_i + c} \left\{ n \int_0^1 r^{n-1}(1-r)^{\sum_i \beta_i + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r \right. \\
&\quad \left. - \left(n + \sum_i \beta_i - 1 \right) \int_0^1 r^{n}(1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r \right\} \\
&\quad \Big/ \left\{ \int_0^1 r^{n}(1-r)^{\sum_k \beta_k + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r \right\}.
\end{aligned}
$$

From Lemma 2 in the Appendix, we have

$$
\begin{aligned}
\psi(n, \beta, c, z) &\geqslant \left(\sum_i \beta_i + c \right) \frac{n + \sum_j \beta_j - 1}{n + \sum_k \beta_k + c} - n \log \frac{n + \sum_i \beta_i + c}{n} \\
&\quad + \frac{(c+1)(1-z) \int_0^1 r^{n+1}(1-r)^{\sum_i \beta_i + c - 1}(1-zr)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^{n}(1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r} \\
&\quad - \frac{n}{n + \sum_i \beta_i + c} \frac{(c+1)(1-z) \int_0^1 r^{n}(1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^{n}(1-r)^{\sum_k \beta_k + c - 1}(1-zr)^{-(c+1)}\,\mathrm{d}r} \\
&= \left(\sum_i \beta_i + c \right) \frac{n + \sum_j \beta_j - 1}{n + \sum_k \beta_k + c} - n \log \frac{n + \sum_i \beta_i + c}{n}
\end{aligned}
$$

$$
+\frac{n}{n+\sum_i \beta_i + c} \frac{(c+1)(1-z)\int_0^1 r^n (1-r)^{\sum_j \beta_j + c - 1}(1-zr)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^n (1-r)^{\sum_k \beta_k + c-1}(1-zr)^{-(c+1)}\,\mathrm{d}r}
\times\left\{\frac{n+\sum_i \beta_i + c}{n}\frac{\int_0^1 r^{n+1}(1-r)^{\sum_j \beta_j + c-1}(1-zr)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^n(1-r)^{\sum_k \beta_k + c - 1}(1-zr)^{-(c+2)}\,\mathrm{d}r} - 1\right\}.
$$

From Lemma 3, we have the inequality

$$
\frac{\int_0^1 r^{n+1}(1-r)^{\sum_i \beta_i + c-1}(1-zr)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^n (1-r)^{\sum_j \beta_j + c-1}(1-zr)^{-(c+2)}\,\mathrm{d}r} \geqslant \frac{B(n+2, \sum_i \beta_i + c)}{B(n+1, \sum_j \beta_j + c)} = \frac{n+1}{n+\sum_j \beta_j + c+1}
$$

by putting $h(r) = (1-zr)^{-(c+2)}$. Therefore,

$$
\psi(n,\beta,c,z) \geqslant \left(\sum_i \beta_i + c\right)\frac{n+\sum_j \beta_j - 1}{n+\sum_k \beta_k + c} - n\log\frac{n+\sum_i \beta_i + c}{n}
+\frac{n}{n+\sum_i \beta_i + c}\frac{(c+1)(1-z)\int_0^1 r^n(1-r)^{\sum_j \beta_j + c-1}(1-zr)^{-(c+2)}\,\mathrm{d}r}{\int_0^1 r^n(1-r)^{\sum_k \beta_k + c-1}(1-zr)^{-(c+1)}\,\mathrm{d}r}
\times\left\{\frac{n+\sum_i \beta_i + c}{n}\frac{n+1}{n+\sum_j \beta_j + c+1} - 1\right\}.
$$

Since $c+1>0$, $0 \leqslant z < 1$, and $\{(n+\sum_i \beta_i + c)/n\}/\{(n+\sum_i \beta_i + c+1)/(n+1)\} - 1 > 0$ when $\sum_i \beta_i + c > 0$, we have

$$
\psi(n,\beta,c,z) > \left(\sum_i \beta_i + c\right)\frac{n+\sum_j \beta_j - 1}{n+\sum_k \beta_k + c} - n\log\frac{n+\sum_i \beta_i + c}{n}.
$$

Since the inequality $\log x \leqslant (1/2)(x - 1/x)$ holds for $x \geqslant 1$,

$$
\begin{aligned}
\psi(n,\beta,c,z) &> \left(\sum_i \beta_i + c\right)\frac{n+\sum_j \beta_j - 1}{n+\sum_k \beta_k + c} - \frac{n}{2}\left(\frac{n+\sum_i \beta_i + c}{n} - \frac{n}{n+\sum_i \beta_i + c}\right)\\
&= \left(\sum_i \beta_i + c\right)\frac{\sum_j \beta_j - c - 2}{2(n+\sum_k \beta_k + c)}. \qquad (6)
\end{aligned}
$$

Since $\sum_i \beta_i + c > 0$ and $\sum_i \beta_i - c - 2 \geqslant 0$ by the assumption, (6) is nonnegative. Thus we have the desired result. □

Corollary 1. *Suppose $d \geqslant 3$. Let b be an arbitrary positive number. If $-1 < c \leqslant d/2 - 2$ and $a \geqslant 1/\kappa$, then the Bayesian predictive density $p_{\pi_{\beta,c,\kappa}}(y|x)$ based on the prior $\pi_{\beta=(1/2,\ldots,1/2),c,\kappa}(\lambda)$ dominates the Bayesian predictive density $p_{\pi_{\mathrm{J}}}(y|x)$ based on the Jeffreys prior $\pi_{\mathrm{J}}(\lambda) = \prod_{i=1}^d \lambda_i^{-1/2}$.*

Proof. The Jeffreys prior density is equal to $\tilde{\pi}_{\beta=(\frac{1}{2},\ldots,\frac{1}{2})}(\lambda)$. Thus the condition $\sum_i \beta_i - c - 2 \geqslant 0$ in Theorem 2 implies $c \leqslant d/2 - 2$. The inequality $\sum_i \beta_i + c = d/2 + c > 0$ is satisfied if $c > -1$. □

Since $\pi_{\beta,c,\kappa}(\lambda)$ is proper when $c > 0$, there exist proper priors dominating the Jeffreys prior when $d \geqslant 5$.

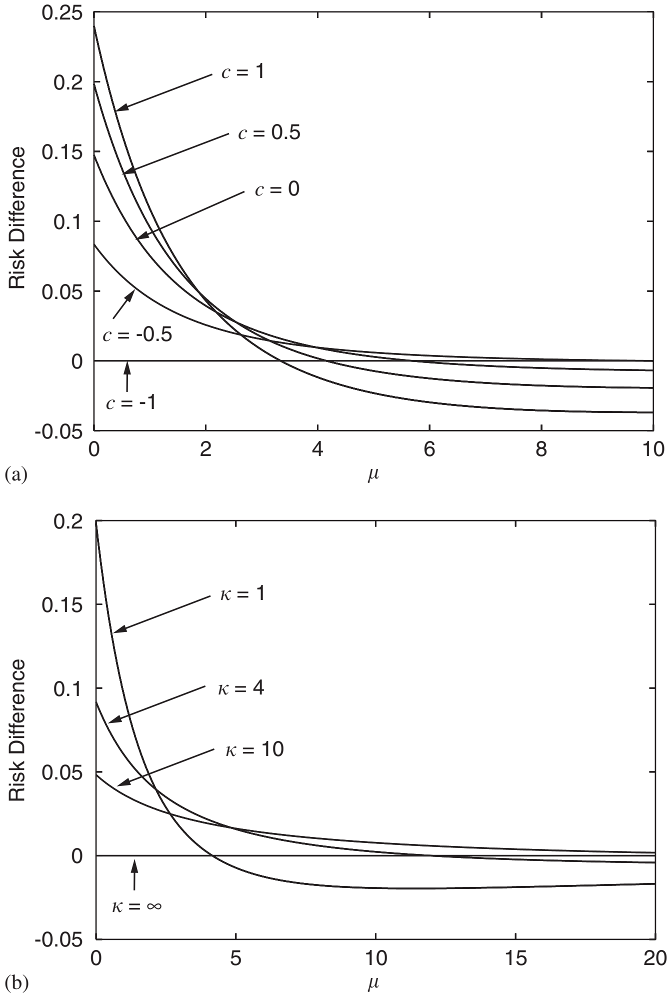

Fig. 1. The difference between the expected divergences $E[D(p(y|\lambda), p_{\pi_{\mathrm{S}}}(y|x))|\lambda] - E[D(p(y|\lambda), p_{\pi_{\beta=(1/2,\ldots,1/2),c,\kappa}}(y|x))|\lambda]$, which depends on λ only through $\mu = \lambda_1 + \lambda_2 + \cdots + \lambda_d$. (a) $d = 6$, $a = 1$, $b = 1$, $\kappa = 1$, and $c = -1, -0.5, 0, 0.5$, and 1. (b) $d = 6$, $c = 0.5$, $a = 1$, $b = 1$, and $\kappa = 1, 4, 10$, and ∞.

Fig. 1 shows the difference between the risk of $p_{\pi_{\beta,c,\kappa}}(y|x)$ and that of $p_{\pi_{\mathrm{S}}}(y|x)$ based on

$$\pi_{\mathrm{S}}(\lambda) = \pi_{\beta=(1/2,\ldots,1/2)}(\lambda) = \frac{1}{(\sum_i \lambda_i)^{\frac{d}{2}-1}(\lambda_1\lambda_2\cdots\lambda_d)^{\frac{1}{2}}}$$

investigated by Komaki [5]. When μ is close to 0, the risk of $p_{\pi_{\beta,c,\kappa}}(y|x)$ is smaller than that of $p_{\pi_{\mathrm{S}}}(y|x)$. When μ is large, the risk of $p_{\pi_{\beta,c,\kappa}}(y|x)$ is larger than that of $p_{\pi_{\mathrm{S}}}(y|x)$. Since $p_{\pi_{\mathrm{S}}}(y|x)$ is admissible [5], the risk of $p_{\pi_{\beta,c,\kappa}}(y|x)$ cannot be smaller than that of $p_{\pi_{\mathrm{S}}}(y|x)$ for every μ. When a prior is proper, it can be easily shown that the corresponding Bayesian predictive density is admissible.

3. Parameter estimation

In this section, we show that the class of prior densities introduced in the previous section is useful also for constructing Bayes estimators. In the case of the independent Poisson observables model, parameter estimation under Kullback–Leibler loss can be regarded as infinitesimal prediction.

There are few studies of estimation under Kullback–Leibler loss $D(p(y|\lambda), p(y|\hat{\lambda}(x)))$ except for the work by Ghosh and Yang [2], which characterized linear admissible estimators of the form $\hat{\lambda}_i = c_i x_i + b_i$, compared with the number of studies based on other loss functions such as squared-error. What is called Stein's loss is the Kullback–Leibler divergence $D(p(y|\hat{\lambda}(x)), p(y|\lambda))$ with the direction opposite to our setting.

First, we consider the model (2). The Kullback–Leibler divergence from the true probability density $p(y|\lambda)$ to a plug-in density $p(y|\hat{\lambda})$ is given by

$$
\begin{aligned}
D(p(y|\lambda), p(y|\hat{\lambda})) &= D(\prod_{i=1}^{d} p(y_i|\lambda_i), \prod_{j=1}^{d} p(y_j|\hat{\lambda}_j)) \\
&= \sum_{i=1}^{d} \left\{ \sum_{y_i=0}^{\infty} \exp(-b\lambda_i) \frac{(b\lambda_i)^{y_i}}{y_i!} \log \frac{\exp(-b\lambda_i)\dfrac{(b\lambda_i)^{y_i}}{y_i!}}{\exp(-b\hat{\lambda}_i)\dfrac{(b\hat{\lambda}_i)^{y_i}}{y_i!}} \right\} \\
&= b \sum_{i=1}^{d} \left(\hat{\lambda}_i - \lambda_i + \lambda_i \log \frac{\lambda_i}{\hat{\lambda}_i} \right).
\end{aligned} \tag{7}
$$

Suppose that a prior density $\pi(\lambda)$ is adopted and the observation x is given. Then

$$
\begin{aligned}
&\int D(p(y|\lambda), p(y|\hat{\lambda}(x))) p_\pi(\lambda|x) \,\mathrm{d}\lambda \\
&= b \int \sum_{i=1}^{d} \left\{ \left(\hat{\lambda}_i(x) - \bar{\lambda}_i(x) + \bar{\lambda}_i(x) \log \frac{\bar{\lambda}_i(x)}{\hat{\lambda}_i(x)} \right) \right. \\
&\quad \left. + \left(\bar{\lambda}_i(x) - \lambda_i + \lambda_i \log \frac{\lambda_i}{\bar{\lambda}_i(x)} \right) \right\} p_\pi(\lambda|x) \,\mathrm{d}\lambda,
\end{aligned}
$$

where $\bar{\lambda}(x)$ is the posterior mean based on the prior density $\pi(\lambda)$ and the observation x, is minimized when $\hat{\lambda}(x) = \bar{\lambda}(x)$. Thus, the (possibly generalized) Bayes estimator is the posterior mean given the observation x.

Next, we consider the prediction problem introduced in Section 1 and put $a = 1$ for simplicity and consider the limit $b \to 0$. Then the limit of the risk divided by b can be regarded as the risk for the infinitesimal prediction. Let $\hat{\lambda}^{\pi}(x)$ be the posterior mean, which is the (possibly generalized) Bayes estimator of λ, based on the prior density $\pi(\lambda)$ and the observation x. Then the limit is given by

$$
\lim_{b \to \infty} \frac{1}{b} \sum_{y} p(y|\lambda) \log \frac{p(y|\lambda)}{p_\pi(y|x)}
$$

$$
= \lim_{b \to 0} \frac{1}{b} \sum_y \left(\prod_{i=1}^d \exp(-b\lambda_i) \frac{(b\lambda_i)^{y_i}}{y_i!} \right.
$$

$$
\left. \times \log \frac{\prod_{i=1}^d \exp(-b\lambda_i) \frac{(b\lambda_i)^{y_i}}{y_i!}}{\frac{\int \pi(\bar{\lambda}) \prod_{i=1}^d [\exp\{-(1+b)\bar{\lambda}_i\}\{(1+b)\bar{\lambda}_i\}^{x_i+y_i}] \, \mathrm{d}\bar{\lambda}}{\int \pi(\bar{\lambda}) \prod_{i=1}^d \{\exp(-\bar{\lambda}_i)(\bar{\lambda}_i)^{x_i}\} \, \mathrm{d}\bar{\lambda}} \prod_{j=1}^d \frac{b^{y_j}}{(1+b)^{x_j+y_j} y_j!}} \right)
$$

$$
= \sum_{i=1}^d \left(\hat{\lambda}_i^{\pi}(x) - \lambda_i + \lambda_i \log \frac{\lambda_i}{\hat{\lambda}_i^{\pi}(x)} \right).
$$

The limit coincides with (7) divided by b. Thus, in the case of the independent Poisson observables model, the estimation problem can be regarded as a special case of the prediction problem. Therefore, in the same way as in the proofs of Theorem 2 and Corollary 1, we can prove the following results.

Theorem 3. *Suppose that we have an observation x from the probability density* (1) *with $a = 1$. Let $\pi_{\beta,c,\kappa}(\lambda)$ be a prior density defined by* (3) *with $\sum_{i=1}^d \beta_i + c > 0$, $\kappa > 0$, and $\beta_i > 0$ ($i = 1, 2, \ldots, d$). If $c > -1$, $\sum_{i=1}^d \beta_i - c - 2 \geqslant 0$ and $\kappa \geqslant 1$, the posterior mean $\hat{\lambda}^{\pi_{\beta,c,\kappa}}(x)$ based on the prior $\pi_{\beta,c,\kappa}(\lambda)$ dominates the posterior mean $\hat{\lambda}^{\tilde{\pi}_\beta}(x)$ based on the prior $\tilde{\pi}_\beta(\lambda) := \prod_{i=1}^d \lambda_i^{\beta_i - 1}$ under Kullback–Leibler loss* (7).

Corollary 2. *Suppose $d \geqslant 3$. If $-1 < c \leqslant d/2 - 2$ and $\kappa \geqslant 1$, then the posterior mean based on the prior $\pi_{\beta=(1/2,\ldots,1/2),c,\kappa}(\lambda)$ dominates that based on the Jeffreys prior $\pi_{\mathrm{J}}(\lambda) = \prod_{i=1}^d \lambda_i^{-1/2}$ under Kullback–Leibler loss.*

Thus, when $d \geqslant 5$, there exist proper Bayes estimators dominating the generalized Bayes estimator based on the Jeffreys prior. This result for the independent Poisson observables model corresponds to the result by Strawderman [9] for $N_d(\mu, I)$ ($d \geqslant 5$).

4. Discussion

The (i, j)-component of the Fisher information matrix of the independent Poisson observables model (1) is

$$
g_{ij}(\lambda) := E\left[\frac{\partial}{\partial \lambda_i} \log p(x|\lambda) \frac{\partial}{\partial \lambda_j} \log p(x|\lambda) \middle| \lambda \right] = \begin{cases} \dfrac{1}{\lambda_i} & (\text{for } i = j) \\ 0 & (\text{for } i \neq j), \end{cases} \tag{8}
$$

where, for simplicity, we put $a = 1$.

The ratio of the prior density $\pi_{\beta=(1/2,\ldots,1/2),c,\kappa}(\lambda)$ and the Jeffreys prior $\pi_{\mathrm{J}}(\lambda)$ is

$$\frac{\pi_{\beta=(1/2,\ldots,1/2),c,\kappa}(\lambda)}{\pi_{\mathrm{J}}(\lambda)} = \frac{\left(\prod_{i=1}^{d}\lambda_i^{-1/2}\right)\int_0^\infty \exp\left(-\frac{\sum_{j=1}^{d}\lambda_j}{s}\right) s^{-d/2}(s+\kappa)^{-(c+1)}\,\mathrm{d}s}{\prod_{i=1}^{d}\lambda_i^{-1/2}}$$

$$= \int_0^\infty \frac{1}{s^{d/2}} \exp\left(-\frac{\sum_{j=1}^{d} {x_j}^2}{4s}\right)(s+\kappa)^{-(c+1)}\,\mathrm{d}s, \qquad (9)$$

where $x_i := 2\sqrt{\lambda_i}$ $(i = 1, 2, \ldots, d)$. The function (9) of x coincides with the ratio of Strawderman's prior and the uniform prior for μ of $N_d(\mu, I)$. Here, $1-c$ corresponds to a in [9].

We consider the parameter space $\{\lambda|\lambda \in \mathbb{R}^d, \lambda_i > 0 \text{ for } 1 \leqslant i \leqslant d\}$ of the independent Poisson observables model as a Riemannian manifold endowed with the Fisher metric (8). Then, the map $(\lambda_1, \lambda_2, \ldots, \lambda_d) \mapsto (2\sqrt{\lambda_1}, 2\sqrt{\lambda_2}, \ldots, 2\sqrt{\lambda_d}) =: (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ is an isometry of the model manifold endowed with the Fisher metric onto $\mathbb{R}^{d+} := \{x|x \in \mathbb{R}^d, x_i > 0 \text{ for } 1 \leqslant i \leqslant d\}$ endowed with the usual Euclidean metric. Therefore, we can identify the model manifold with $\mathbb{R}^{d+}$.

When we consider the parameter space $\{\mu|\mu \in \mathbb{R}^d\}$ of the d-dimensional normal model $N_d(\mu, I)$ as a model manifold endowed with the Fisher metric, the model manifold is isometric to the Euclidean space $\mathbb{R}^d$. Therefore, the independent Poisson observables model and the multivariate normal model with known covariance matrix have similar differential geometric properties, and it is natural from an asymptotic viewpoint that similar results hold for the multivariate normal model and the independent Poisson observables model, for details see [6].

The results in the present paper have been proved by showing monotonicity of (4). On the other hand, for the multivariate normal model, George et al. [1] have obtained several handy sufficient conditions under which a Bayesian predictive density based on a prior dominates that based on the Jeffreys prior. Similar sufficient conditions could be useful also for the independent Poisson observables model, and further research is required.

Appendix

We give several lemmas and proofs.

Lemma 1. *Let t be a positive real number, and let $\pi_{\beta,c,\kappa}(\lambda)$ be a prior density defined by* (3) *with $\sum_{i=1}^{d}\beta_i + c > 0$, $\kappa > 0$, and $\beta_i > 0$ $(i = 1, 2, \ldots, d)$. Then,*

$$\int \pi_{\beta,c,\kappa}(\lambda)\prod_{i=1}^{d}\left\{\exp(-t\lambda_i)(t\lambda_i)^{x_i}\right\}\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2\cdots\mathrm{d}\lambda_d$$

$$= t^{-1}\kappa^{-(c+1)}\prod_{i=1}^{d}\Gamma(x_i+\beta_i)\int_0^1 r^{\sum_j x_j}(1-r)^{\sum_k \beta_k + c - 1}\left\{1-\left(1-\frac{1}{t\kappa}\right)r\right\}^{-(c+1)}\mathrm{d}r$$

Proof. We have

$$\int \pi_{\beta,c,\kappa}(\lambda)\prod_{i=1}^{d}\left\{\exp(-t\lambda_i)(t\lambda_i)^{x_i}\right\}\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2\cdots\mathrm{d}\lambda_d$$

$$
\begin{aligned}
&= \int \left(\prod_i \lambda_i^{\beta_i - 1}\right) \int_0^\infty \exp\left(-\frac{\sum_j \lambda_j}{s}\right) s^{-\sum_k \beta_k} (s+\kappa)^{-(c+1)}\, \mathrm{d}s \\
&\quad \times \prod_{l=1}^d \left\{\exp(-t\lambda_l)(t\lambda_l)^{x_l}\right\} \mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_d \\
&= t^c \prod_{i=1}^d \Gamma(x_i + \beta_i) \int_0^\infty \left(1 + \frac{1}{ts}\right)^{-\sum_j (x_j + \beta_j)} (ts)^{-\sum_k \beta_k} (ts + t\kappa)^{-(c+1)}\, \mathrm{d}(ts).
\end{aligned}
$$

By putting $ts = r/(1-r)$, we have

$$
\begin{aligned}
&\int \pi_{\beta,c,\kappa}(\lambda) \prod_{k=1}^d \left\{\exp(-t\lambda_k)(t\lambda_k)^{x_k}\right\} \mathrm{d}\lambda_1\, \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_d \\
&= t^c \prod_{i=1}^d \Gamma(x_i + \beta_i) \int_0^1 r^{\sum_j (x_j + \beta_j)} \left(\frac{r}{1-r}\right)^{-\sum_k \beta_k} \left(\frac{r}{1-r} + t\kappa\right)^{-(c+1)} \frac{1}{(1-r)^2}\, \mathrm{d}r \\
&= t^{-1}\kappa^{-(c+1)} \prod_{i=1}^d \Gamma(x_i + \beta_i) \int_0^1 r^{\sum_k x_k} (1-r)^{\sum_i \beta_i + c - 1} \\
&\quad \times \left\{1 - \left(1 - \frac{1}{t\kappa}\right) r\right\}^{-(c+1)} \mathrm{d}r. \qquad \square
\end{aligned}
$$

The following lemma is equivalent to one of Gauss' recursion formulae for the hypergeometric function.

Lemma 2. *Suppose* $u > 0$, $v > 0$, $z > 0$, *and* $a \in \mathbb{R}$. *Then,*

$$
\begin{aligned}
&u \int_0^1 r^{u-1}(1-r)^{v-1}(1-zr)^{-a}\, \mathrm{d}r - (u+v-a) \int_0^1 r^u (1-r)^{v-1} (1-zr)^{-a}\, \mathrm{d}r \\
&-a(1-z) \int_0^1 r^u (1-r)^{v-1}(1-zr)^{-(a+1)}\, \mathrm{d}r = 0.
\end{aligned}
$$

Proof. By partial integration, we have

$$
\begin{aligned}
&az \int_0^1 r^u (1-r)^v (1-zr)^{-(a+1)}\, \mathrm{d}r \\
&= -u \int_0^1 r^{u-1}(1-r)^v (1-zr)^{-a}\, \mathrm{d}r + v \int_0^1 r^u (1-r)^{v-1}(1-zr)^{-a}\, \mathrm{d}r.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&u \int_0^1 r^{u-1}(1-r)^v (1-zr)^{-a}\, \mathrm{d}r - v \int_0^1 r^u (1-r)^{v-1} (1-zr)^{-a}\, \mathrm{d}r \\
&+a \int_0^1 r^u (1-r)^{v-1}(1-zr)^{-(a+1)}\, \mathrm{d}r - az \int_0^1 r^{u+1}(1-r)^{v-1}(1-zr)^{-(a+1)}\, \mathrm{d}r \\
&+(-a+az) \int_0^1 r^u (1-r)^{v-1} (1-zr)^{-(a+1)}\, \mathrm{d}r = 0.
\end{aligned}
$$

Therefore,

$$u\int_0^1 r^{u-1}(1-r)^{v-1}(1-zr)^{-a}\,\mathrm{d}r-(u+v-a)\int_0^1 r^{u}(1-r)^{v-1}(1-zr)^{-a}\,\mathrm{d}r$$
$$-a(1-z)\int_0^1 r^{u}(1-r)^{v-1}(1-zr)^{-(a+1)}\,\mathrm{d}r=0. \qquad \square$$

Lemma 3. *Suppose that $h(r)$ is a bounded nondecreasing function of $r \in [0, 1]$. Let u and v be positive real numbers. Then*,

$$\int_0^1 \frac{r^u(1-r)^{v-1}}{B(u+1,v)}h(r)\,\mathrm{d}r \geqslant \int_0^1 \frac{r^{u-1}(1-r)^{v-1}}{B(u,v)}h(r)\,\mathrm{d}r$$

Proof. Let $f(x)$ and $g(x)$ be nondecreasing functions of $x \in \mathbb{R}$. Suppose that X is a real random variable and the variances of $f(X)$ and $g(X)$ exist. Let $\bar{f} = E[f(X)]$. Since $f(x)$ is nondecreasing, there exists c such that $f(x) \geqslant \bar{f}$ for $x \geqslant c$ and $f(x) \leqslant \bar{f}$ for $x \leqslant c$. Since $E[f(X)g(X)] - E[f(X)]E[g(X)] = E[(f(X)-\bar{f})g(X)] = E[(f(X)-\bar{f})(g(X)-g(c))]$ and $(f(x)-\bar{f})(g(x)-g(c)) \geqslant 0$ for all x, we have $E[f(X)g(X)] \geqslant E[f(X)]E[g(X)]$. Let $f(r) := h(r)$, $g(r) := r$, and $E[f(X)] := \int \frac{r^{u-1}(1-r)^{v-1}}{B(u,v)} f(r)\,\mathrm{d}r$. Then, $E[f(X)g(X)]/E[g(X)] \geqslant E[f(X)]$ means the desired inequality. $\square$

References

[1] E.I. George, F. Liang, X. Xu, Improved minimax prediction under Kullback–Leibler loss, Ann. Statist. 34 (2006), to appear.

[2] M. Ghosh, M.-C. Yang, Simultaneous estimation of Poisson means under entropy loss, Ann. Statist. 16 (1988) 278–291.

[3] H. Jeffreys, Theory of Probability, third ed., Oxford University Press, Oxford, 1961.

[4] F. Komaki, A shrinkage predictive distribution for multivariate normal observables, Biometrika 88 (2001) 859–864.

[5] F. Komaki, Simultaneous prediction of independent Poisson observables, Ann. Statist. 32 (2004) 1744–1769.

[6] F. Komaki, Shrinkage priors for Bayesian prediction, Ann. Statist. 34 (2006), to appear.

[7] F. Liang, Exact minimax procedures for predictive density estimation and data compression, Ph.D. Thesis, Department of Statistics, Yale University, 2002.

[8] C. Stein, Estimation of the mean of a multivariate normal distribution, in: J. Hájek (Ed.), Proceedings of the Prague Symposium on Asymptotic Statistics, Universita Karlova, Prague, 1974, pp. 345–381.

[9] W.E. Strawderman, Proper bayes minimax estimators of multivariate normal mean, Ann. Math. Statist. 42 (1971) 385–388.